

Erschienen in: Engelberg, Stefan/Holler, Anke/Proost, Kristel (Hrsg.): Sprachliches Wissen zwischen Lexikon und Grammatik. Berlin/Boston: de Gruyter, 2011. (Jahrbuch des Instituts für Deutsche Sprache 2010), S. 559-572.

Carolin Müller-Spitzer / Stefan Engelberg

Elektronische Lexikografie zwischen Grammatik und Lexikon

Abstract

Der Artikel stellt die Projekte vor, die sich im Rahmen der Projektmesse zur „Elektronischen Lexikografie“ präsentiert haben. Diese Messe wurde begleitend zur 46. Jahrestagung des Instituts für Deutsche Sprache veranstaltet. Es wird in diesem Beitrag auf der Basis der Messepräsentationen dargelegt, inwiefern Entwicklungen der Korpuslexikografie und der Internetlexikografie die lexikografische Erfassung syntagmatischer Aspekte des deutschen Wortschatzes befördern und welche lexikografischen Internetressourcen dazu verfügbar sind.

1. Korpuslexikografie und Internetlexikografie

Die Entwicklung der Lexikografie in der vergangenen Dekade ist zum einen gekennzeichnet durch den Einsatz immer größerer elektronischer Textkorpora und der Verfeinerung der Methoden zu ihrer automatischen Analyse (Korpuslexikografie) und zum anderen durch eine Hinwendung der elektronischen Lexikografie zum Internet (Internetlexikografie).

Sowohl die Fortschritte der Korpuslexikografie wie auch die neuen medienspezifischen Möglichkeiten der Internetlexikografie haben dazu geführt, dass Phänomene im Übergangsbereich zwischen Lexikon und Grammatik in den Fokus der lexikografischen Theorie und Praxis gerückt sind. Korpusanalysemethoden wie Kookkurrenzanalysen, N-Gramm-Analysen und Frequenzanalysen gehören mittlerweile zum Standard der lexikografischen Praxis (vgl. z.B. Klosa 2007; Prinsloo 2009). Kookkurrenzanalysen berechnen das überzufällig häufige gemeinsame Auftreten von Wörtern und ermöglichen so die (semi-)automatische Ermittlung von Kollokationen, Redewendungen, Funktionsverbgefügen und anderen Arten von Phrasemen. N-Gramm-Analysen ermitteln die Häufigkeit von Wortkombinationen und damit die Gebräuchlichkeit von Mehrwortausdrücken. Über andere Arten von Frequenzanalysen wird die Häufigkeit lexikalischer Ausdrücke in bestimmten Typen von Konstruktionen errechnet. All diese Verfahren haben zu der Erkenntnis geführt, dass die Anzahl von Mehrwortausdrücken, die über nicht regelhafte formale oder nicht-kompositionelle inhaltliche Idiosynkrasien verfügen oder die aufgrund von Frequenzauffälligkeiten einen sprachsystematisch nicht zu erklärenden Usualisierungsgrad aufweisen, um ein Vielfaches größer ist als in präkorpuslinguistischen Zeiten angenommen wurde.

Die Hinwendung der Lexikografen zur elektronischen und neuerdings besonders zur Internetlexikografie hat wiederum die für die Printlexikografie typischen Platzbeschränkungen hinfällig werden lassen, die eine ausführliche Darstellung der Syntagmatik des Wortschatzes oft verhindern. Die großen gegenwartssprachlichen internetlexikografischen Projekte wie das „Digitale Wörterbuch der Deutschen Sprache“ (siehe unten Projekt 4) und „*lexiko*“ (siehe unten Projekt 14) sowie der vorwiegend automatisch erstellte „Deutsche Wortschatz“ (siehe unten Projekt 3) zeichnen sich insofern durch eine besondere Berücksichtigung syntagmatischer Aspekte von Wörtern aus.

2. Projektmesse „Elektronische Lexikografie“

Auf der 46. Jahrestagung des Instituts für Deutsche Sprache, die sich mit methodisch-empirischen und theoretischen Fragen des sprachlichen Wissens zwischen Lexikon und Grammatik befasste, wurde im Zusammenhang mit den oben dargestellten Entwicklungen in der Lexikografie eine von Carolin Müller-Spitzer organisierte Projektmesse „Elektronische Lexikografie“ veranstaltet (vgl. auch Müller-Spitzer/Vachková 2010): Die Projektmesse war dabei nicht beschränkt auf den oben beschriebenen zentralen Themenbereich der Syntagmatik, beinhaltete aber eine Vielzahl von Präsentationen, die in der einen oder anderen Weise dafür relevant sind. Insgesamt stellten sich 16 Projekte vor, von denen einige spezifisch dem Bereich der Syntagmatik zuzuordnen sind (Projekt 11 „E-VALBU“, Projekt 12 „Erlanger Valency Patternbank“, Projekt 13 „Usuelle Wortverbindungen“ bzw. „SprichWort“), die zum Teil aber auch in den Bereich der automatischen Sprachverarbeitung gehören wie „GermaNet“ (Projekt 6, 7) oder „SALSA“ (Projekt 8), internationale internetlexikografische Projekte wie das „Algemeen Nederlands Wordenboek“ (Projekt 1) oder das „Große akademische Wörterbuch Deutsch-Tschechisch“ (Projekt 2), Forschungsprojekte im Bereich der Internetlexikografie wie „Kognitiv Saliente Semantische Relationen“ (Projekt 10) oder „Benutzeradaptive Zugänge und Vernetzungen in *lexiko*“ (Projekt 14). Darüber hinaus waren Projekte eingeladen, die entweder im Bereich der Korpuserstellung Besonderes zu berichten hatten, wie das Projekt 5 im Bereich der Deutschen Gebärdensprache, oder hinsichtlich der Darstellung innovative Formen wählen, wie das „Kicktionary“ (Projekt 9). Die Messe diente damit zum einen dazu, den Tagungsteilnehmern Projekte aus der Praxis zu zeigen, die einige der in dem Programm der Jahrestagung theoretisch beleuchteten Aspekte lexikografisch umsetzen, und zum anderen dazu, den Austausch der Forscher und Lexikografen aus dem Bereich Internetlexikografie untereinander zu fördern.

- **Projekt 1 (*Algemeen Nederlands Wordenboek*):** Das ANW ist die Neuerarbeitung eines allgemeinen, einsprachigen Wörterbuchs gezielt für das elektronische Medium. Das Besondere hier ist u.a. der Ansatz einer formalisierten Bedeutungsbeschreibung durch so genannte Semagramme, die interessante und neuartige erweiterte Suchen im elektronischen Wörterbuch ermöglichen sollen. Semagramme sind semantische Merkmalskarten, in denen eine Vielzahl semantischer Eigenschaften zu einer Lesart möglichst formalisiert in festen Kategorien festgehalten werden (Moerdijk 2008; Moerdijk/Tiberius/Niestadt 2008).
- **Projekt 2 (*Das Große akademische Wörterbuch Deutsch-Tschechisch*):** In dem Prager Wörterbuchprojekt werden Kookurrenzanalysen sowie darauf basierende „Self-Organizing Lexical Feature Maps“ als korpuslinguistische Methoden eingesetzt, insbesondere zur Ermittlung von semantischen Relationen (Vachková/Belica 2009).
- **Projekt 3 (*Deutscher Wortschatz der Universität Leipzig*):** Der induktiv aus Korpora gewonnene „Deutsche Wortschatz“ zeichnet sich unter anderem durch automatisch berechnete Kollokationen und Hyponymie-Relationen aus (Langanke 2004).
- **Projekt 4 (*Digitales Wörterbuch zur deutschen Sprache – DWDS*):** Das DWDS basiert auf dem „Wörterbuch der deutschen Gegenwartssprache“ (Klappenbach/Steinitz (Hg.) 1967–1977), das mittlerweile durch eine Vielzahl weiterer Angaben und Funktionen erweitert wurde. Unter anderem können automatisch berechnete Kollokationen – auch grafisch visualisiert – als Ergänzung zu den Wortartikeln rezipiert werden (Klein 2004).
- **Projekt 5 (*Entwicklung eines korpusbasierten, elektronischen Wörterbuchs Deutsche Gebärdensprache – Deutsch*):** Das „Wörterbuch Deutsche Gebärdensprache“ ist u.a. mit den Fragen konfrontiert, wie man ein Korpus zur Gebärdensprache aufbaut, welche Datenschutzprobleme dabei auftreten und wie man mit gebärdensprachlichen Varietäten umgeht.
- **Projekt 6 (*GermaNet – ein lexikalisch semantisches Wortnetz*):** *GermaNet* ist ein computerlexikografisch ausgerichtetes Projekt, welches eher die automatische Sprachverarbeitung als menschliche Nutzer anvisiert. GermaNet fügt sich in die Reihe nationaler WordNet-Vorhaben ein (vgl. Fellbaum in diesem Band) und ist mittlerweile in das EuroWordNet integriert. Es fokussiert wie alle WordNet-Projekte auf die lexikalische Vernetzung durch semantische Relationen (Kunze/Lemnitzer 2010).
- **Projekt 7 (*HyTex – Werkzeuge zur Exploration von GermaNet*):** Im Dortmunder „HyTex“-Projekt werden Werkzeuge zur Exploration von GermaNet entwickelt, z.B. eine grafische Sicht auf das lexikalische Netz. Da für GermaNet im akademischen Kontext mittlerweile kostenfrei Lizenzen vergeben werden, gewinnen solche Werkzeuge an Bedeutung.

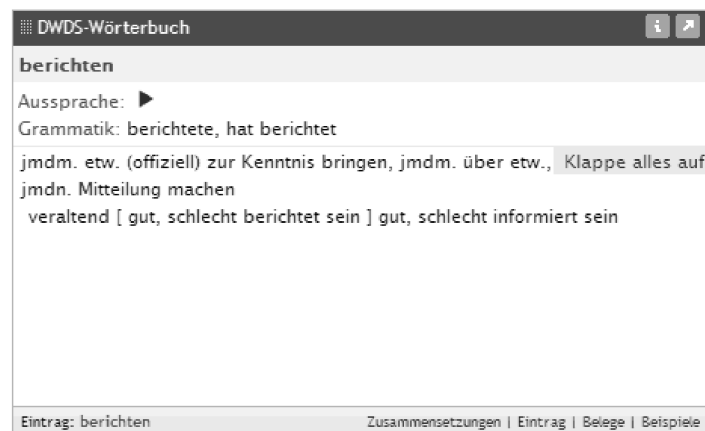
- **Projekt 8 (*Saarbrücken Lexical Semantics Acquisition Project – SALSA*):** Ziel des „SALSA“-Projektes (Burchardt et al. 2009) ist es, angeschlossen an die internationalen FrameNet-Projekte (vgl. Ruppenhofer et al. 2010), ein großes framebasiertes Lexikon für das Deutsche korpusbasiert zu erstellen, das auf semantisch und syntaktisch detailliert annotierte Sätze zugreift. Im Gegensatz zu GermaNet basiert es also auf syntagmatischen und nicht auf paradigmatischen Relationen. Auch SALSA ist primär als Ressource für linguistische und computerlinguistische Forschungen gedacht.
- **Projekt 9 (*Kicktionary*):** Auch das multilinguale elektronische Wörterbuch zur Fußballsprache beruht auf einem framebasierten semantischen Ansatz. Es ermöglicht etwa einen Blick auf das sprachliche Inventar unterschiedlicher „Scenes“ wie z.B. „Shot“, den Schuss. Die einzelnen lexikalischen Einheiten werden nach Frames eingeordnet und sind über Konzepthierarchien zugreifbar (Schmidt 2009).
- **Projekt 10 (*Kognitiv Saliente Semantische Relationen*):** Im Kontext des „ELDIT“-Projektes (Abel et al. 2005; Abel 2008), eines digitalen bilingualen Online-Wörterbuchs Deutsch-Italienisch, ist die Dissertation von Gerald Kremer zu „Kognitiv salienten semantischen Relationen“ angesiedelt. In dem Projekt geht es um Hilfen zur Erweiterung von „ELDIT“ im Bereich der Anzeige von semantischen Relationen zu einem gesuchten Lexikoneintrag in der Lerner-Spache.
- **Projekt 11 (*E-VALBU*):** Das am IDS beheimatete Projekt „E-VALBU“ ist eine für das elektronische Medium aufbereitete Version des „Valenzwörterbuchs deutscher Verben“ (Schumacher et al. 2004). Das auch auf den DaF-Bereich ausgerichtete Wörterbuch zeichnet sich in der Internet-Version durch viele erweiterte Suchmöglichkeiten aus (Kubczak 2009).
- **Projekt 12 (*Erlanger Valency Patternbank*):** In der „Erlanger Valency Patternbank“, die auf dem „Valency Dictionary of English“ (Herbst et al. 2004) basiert, können lexikalische Einheiten nach unterschiedlichen Kriterien durchsucht und sortiert werden; z.B. sind einzelne Muster („Patterns“) über Listen von „Pattern elements“ wie z.B. „about_NP_V-ing“ geordnet und die entsprechenden Verben somit als Menge zugreifbar.
- **Projekte 13a (*Usuelle Wortverbindungen*) und 13b (*Sprich Wort*):** Die beiden in der Abteilung Lexik am Institut für Deutsche Sprache angesiedelten Projekte befassen sich mit Mehrwortausdrücken. In dem Wörterbuch zu „Usuellen Wortverbindungen“ werden Mehrwortausdrücke korpusbasiert in ihrer Varianz dargestellt (Steyer 2009). In dem daran angebundenes EU-Projekt „SprichWort“ wird eine mehrsprachige Sprichwort-Datenbank erstellt.

- **Projekt 14 (*ellexiko*):** In „*ellexiko*“ – einem am Institut für Deutsche Sprache beheimateten Onlinewörterbuch zum Gegenwartsdeutschen – werden alle Artikel neu auf Basis des *ellexiko*-Korpus und gezielt für das elektronische Medium erarbeitet. Besonders interessant für den Kontext der Jahrestagung sind die umfangreichen Angaben zu „Lexikalischen Mitspielern“ sowie zu „Typischen Verwendungen“ (Storjohann/Möhrs 2007).
- **Projekt 15 (*Benutzeradaptive Zugänge und Vernetzungen in ellexiko*):** Einer der drei Schwerpunkte dieses Projektes ist die Benutzungsforschung in Bezug auf Online-Wörterbücher. Im zweiten Schwerpunkt geht es um die Kodierung und Verwaltung sowie um Möglichkeiten der Darstellung von Vernetzungen im lexikografischen Kontext. Der dritte Schwerpunkt hat die automatische Gewinnung morphologischer Angaben für „*ellexiko*“ zum Gegenstand.
- **Projekt 16 (*Online-Wortschatz-Informationssystem Deutsch – OWID*):** Das „Online-Wortschatz-Informationssystem Deutsch“ ist das lexikografische Internet-Portal des Instituts für Deutsche Sprache (Müller-Spitzer 2010; Engelberg/Klosa/Müller-Spitzer 2009). Hier werden alle internetlexikografischen Arbeiten des IDS in einem Portal zusammengeführt. Insbesondere die dort enthaltenen Wörterbücher „*ellexiko*“, „Usuelle Wortverbindungen“ und in naher Zukunft auch „E-VALBU“, das Sprichwörterbuch und das „Handbuch deutscher Kommunikationsverben“ befassen sich mit Phänomenen im Bereich des Tagungsthemas.

3. Internet-Ressourcen zur Syntagmatik des Lexikons der deutschen Gegenwartssprache

Der obige Überblick zeigt, dass viele der auf der Tagung angesprochenen Phänomenbereiche auch zum Gegenstand lexikografischer Projekte geworden sind, insbesondere Kollokationen, feste Wendungen, Valenzen und Argumentstrukturen. Auch andere, nicht auf der Tagung diskutierte Phänomene wie Redewendungen und Sprichwörter sind mittlerweile über elektronische Ressourcen zu erschließen (Projekt 13).

Die am häufigsten diskutierten Phänomene in dem vorliegenden Band sind solche im Umfeld von Valenz und Argumentstruktur. Das sind Phänomene, die gewöhnlich in Wörterbüchern, die Verben verzeichnen, auch als Angabeklassen repräsentiert sind. Große allgemeinsprachliche Internetwörterbücher zum Gegenwartsdeutschen bieten daher im Allgemeinen auf die eine oder andere Weise Informationen zur Rektion bzw. Valenz von Wörtern (siehe Abbildungen 1 und 2).

Abb. 1: Satzbaupläne zu *berichten* in „elexiko“Abb. 2: Valenzangaben zu *berichten* im „DWDS“

Dabei ist es aber für externe Benutzer nicht möglich, Suchen nach Verben mit bestimmten Valenzen/Satzbauplänen/Argumentstrukturen durchzuführen. Dies können etwa in „elexiko“ nur die Lexikografen über einen internen Zugriff. Diese erweiterten Suchen sollen allerdings auch für externe Benutzer deutlich ausgebaut werden. Das „E-VALBU“ (Projekt 11) erlaubt dagegen jetzt schon Suchen über bestimmte formale Valenzmuster. In Abbil-

Abbildung 3 zeigt die Suchmaske und Ergebnisdarstellung zu einer Abfrage nach allen Verben, die außer dem Subjektargument eine NP im Dativ und eine im Akkusativ fordern.

Komplemente:	<input checked="" type="checkbox"/> K _{sub} <input checked="" type="checkbox"/> K _{akk} <input type="checkbox"/> K _{akk2} <input type="checkbox"/> K _{gen} <input checked="" type="checkbox"/> K _{dat} <input type="checkbox"/> K _{pp} <input type="checkbox"/> K _{pp2} <input type="checkbox"/> K _{adv} <input type="checkbox"/> K _{adv2} <input type="checkbox"/> K _{pre} <input type="checkbox"/> K _{itr}	Ergebnisliste:
Satzbauplan:	beliebig	abgeben 4
Passiv:	beliebig	abnehmen 7
Pertinenzelemente:	beliebig	abnehmen 8
Volltextsuche:		abnehmen 9
<input type="button" value="Suche starten"/>	<input type="button" value="Formular leeren"/>	abnehmen 10
Bitte per Mausklick einen Eintrag aus der Ergebnisliste auswählen!		abnehmen 11
		abnehmen 12
		anbieten 1
		anbieten 2
		anbieten 3
		anbieten 4
		antehen 2
		antworten 1
		anziehen 2

Abb. 3: Abfrage zu dreistelligen Nominativ-Akkusativ-Dativ-Verben im „E-VALBU“

Nicht valenz-, sondern frametheoretisch orientiert ist das Lexikon syntagmatischer Angaben zur Syntax und Semantik lexikalischer Einheiten des Deutschen, das in dem Projekt SALSA erstellt wird. Es basiert auf einem detailliert phrasenstrukturell und framesemantisch annotierten Korpus (Projekt 8, siehe Abbildung 4).

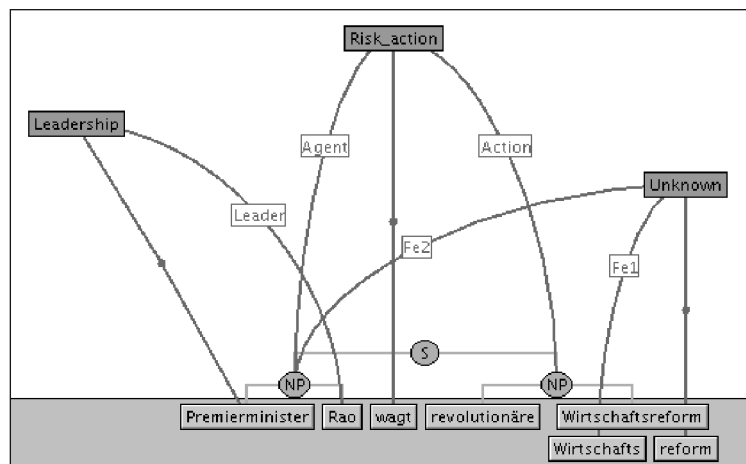


Abb. 4: Phrasenstrukturell und framesemantisch annotierter Satz im SALSA/TIGER-Korpus (www.coli.uni-saarland.de/projects/salsa/gold/page.php?id=s52 (Stand: 24.08.2010))

Im Bereich der lexikografischen Erschließung von Phänomenen im Umfeld von Valenz und Argumentstruktur gibt es allerdings durchaus noch Desiderata. So enthalten allgemeine, einsprachige Internetwörterbücher zum Gegenwartsdeutschen keine Angaben zur Inhaltsvalenz (semantische Rollen, Selektionsrestriktionen), die als Suchkriterien fungieren können. Angaben zu lexikalischen Mitspielern finden sich allerdings in „*lexiko*“. Sie werden dabei aber nicht zu Rollen oder Selektionsrestriktionen generalisiert.

Insbesondere aber sind solche Argumentstrukturmuster, die mit stabilen Bedeutungen einhergehen (Resultativkonstruktionen, Mittelkonstruktionen, Applikativkonstruktionen und viele mehr) nicht nur nicht suchbar, sondern gewöhnlich bezüglich ihres Vorkommens mit bestimmten Verben in Wörterbüchern gar nicht verzeichnet, und das gilt nicht nur für Wörterbücher des Internets (vgl. Engelberg 2010). E-VALBU ermöglicht aber immerhin schon das Suchen nach Verben, die in bestimmten Passiv- oder Perinenzkonstruktionen auftreten.

Eine Ressource, die Argumentstrukturmuster bzw. -konstruktionen im Zusammenspiel mit idiosynkratischen Eigenschaften von Verben erfasst, etwa auf der Basis von empirischen Verfahren wie sie von Stefanowitsch und Gries (Gries/Stefanowitsch 2010; Stefanowitsch in diesem Band) angewendet werden, bleibt allerdings ein Desideratum.

Die Lexikografie von Phrasemen, also etwa Kollokationen, festen Wortverbindungen, Funktionsverbgefügen, Redewendungen und Sprichwörtern, hat von der Entwicklung korpuslinguistischer Methoden schon deutlich mehr profitiert als der Argumentstrukturbereich. Insbesondere die Kookkurrenzanalyse hat dieses Gebiet vorangebracht. Lexikografische Informationssysteme präsentieren entweder vollautomatisch gewonnene typische Wortverbindungen wie im „Deutschen Wortschatz“ (Projekt 3; Abbildung 5) oder lexikografisch überarbeitete, semi-automatisch gewonnene Verbindungen („*lexiko*“, Projekt 14; Abbildung 6).

Signifikante Kookkurrenzen für trinken:

Bier (3505), essen (3098), Kaffee (2522), Alkohol (1405), Tee (1129), Wein (1026), Wasser (1022), zu (828), : (803), rauchen (740), und (617), viel (537), sitzen (487), Liter (423), oder (364), Glas (351), man (332), sie (325), gehen (320), Cola (317), Bierchen (304), Flasche (290), Rotwein (282), etwas (278), mal (267), wir (242), Sie (232), Tasse (230), Wodka (217), Flüssigkeit (187), ich (182), Mineralwasser (181), Sekt (179), Milch (166), Leute (163), Männer (161), Schnaps (160), reden (157), abends (156), Durst (155), dann (155), kann (146), predigen (144), tanzen (141), was (135), einen (134), darf (133), dürfen (121), Kneipe (121), täglich (120), Wenn (116), Bar (115), Tag (115), Viel (113), rät (110), Glühwein (110), Menschen (110), Cocktails (109), Schluck (109), keinen (107), Champagner (104), Espresso (103), Cappuccino (102), Café (100), lieber (99), ein (97), Leitungswasser (95), Gläschen (95), : (94), gerne (93), schlafen (92), Hitze (92), quatschen (91), Essen (89), sollte (88), Getränke (87), Frauen (87), mehr (86), Kölsch (85), gern (85)

Mehrvortkookkurrenzen:

Kaffee trinken (2467), viel trinken (761), Tee trinken (502), zu viel (154), am Tag (139), zu wenig (126), Abwarten und Tee trinken (113), in Ruhe (94), ein Glas Wein (81), abwarten und Tee trinken (80), unter den Tisch trinken (80)

Signifikante linke Nachbarn von trinken:

zu (3252), Bier (2320), Kaffee (2171), Alkohol (991), Wein (625), Tee (604), Wasser (514), viel (366), und (327), Viel (236), Sie (187), Bierchen (180), Rotwein (174), Espresso (103), etwas (101), Champagner (101), Schnaps (96), was (90), Cola (88), mehr (84), Wodka (81), Wir (80), Cappuccino (78), am Tag (75), ein Glas Wein (74), Sekt (73), genügend (68), noch ein Bier (65), zu viel (64), eine Tasse Kaffee (58), wir (57), Kakao (56), Leute (56), ein Glas Wasser (56), Tag (55), Milch (53), um zu (50), Kölsch (50), Viertel (49), Glühwein (48), Liter (47), eine Tasse Tee (46), noch etwas (43), Warmes (43), Deutschen (42), Durst (41), Saft (40), Caipirinha (40), Nektar (38), Apfelwein (38), Gin (38), einen (37), ausreichend (37), pro Tag (37), genug (36), Männer (36), Biere (36), Strohhalm (35), im Freien (35), Pils (34), regelmäßig (34), heiße Schokolade (33), Pilger (33), Cocktails (31), so viel (31), täglich (30), Ebbelwei (30), Obstler (30), zu wenig (29), Heißes (28), Britinnen (27), Apfelschorle (26), Freien (26), Schlafengehen (26), oder (26), Flüssigkeit (25), Merlot (25), Tisch (24), Cognac (24), Alcopops (23)

Abb. 5: Verschiedene Angaben zu Kookkurrenzen zu *trinken* im „Deutschen Wortschatz“

nachweisen
Lesart: 'beweisen'

zur Übersichtsseite Lesarten im Überblick

Bedeutungs-erläuterung Semantische Umgebung **Typische Verwendungen** Sinnverwandte Wörter Besonderheiten des Gebrauchs Grammatik

1 Typische Verwendungen

nachweisen in Infinitiv-Verwendungen verbergen x

- verbotene Substanzen nachweisen
- ihre Unschuld nachweisen
- einen Wohnsitz nachweisen
- vor Gericht nachweisen
- ein Verschulden ist nachzuweisen
- [...] ist kaum mehr nachzuweisen
- [z. B. *ein Betrug*] ist schwer nachzuweisen
- [Person] ist keine Schuld nachzuweisen
- es wird schwierig sein [z. B. *Bestechung*] nachzuweisen

nachweisen in finiten Verwendungen verbergen x

- [Person] muss nachweisen, dass [...]
- die Bewerber müssen eine möglichst mehrjährige Berufserfahrung nachweisen
- eindeutig nachgewiesen sei [z. B. *der negative Einfluß von Zigarettenrauch*]
- dem Angeklagten wurde nachgewiesen, [...]
- die Einnahme verbotener Substanzen wurde nachgewiesen
- [Person] wurde die Einnahme von [z. B. *Ephedrin*] nachgewiesen
- es wurden Spuren von [z. B. *Cannabis*] nachgewiesen
- [z. B. *ein Fremdverschulden*] konnte nicht nachgewiesen werden
- eine Tatbeteiligung konnte nicht nachgewiesen werden
- bei [Person] konnten Antikörper im Blut nachgewiesen werden
- in der A-Probe konnten keine Dopingmittel nachgewiesen werden

Abb. 6: Typische Verwendungen des Verbs *nachweisen* in „elexiko“

Dabei werden die Darstellungen häufiger durch netzbasierte oder typografisch gestaltete grafische Darstellungen ergänzt (siehe Abbildungen 7 und 8).

Mittlerweile sind aber auch spezifische neue Ressourcen für Phraseme ins Netz gegangen. Die im Aufbau befindliche Internetressource „Wortverbindungsfelder Online“ des Projekts „Usuelle Wortverbindungen“ (Projekt 13) dokumentiert detailliert auf der Basis von Korpusanalysen den Gebrauch und die Varianz von Wortverbindungen (siehe Abbildungen 9 und 10).

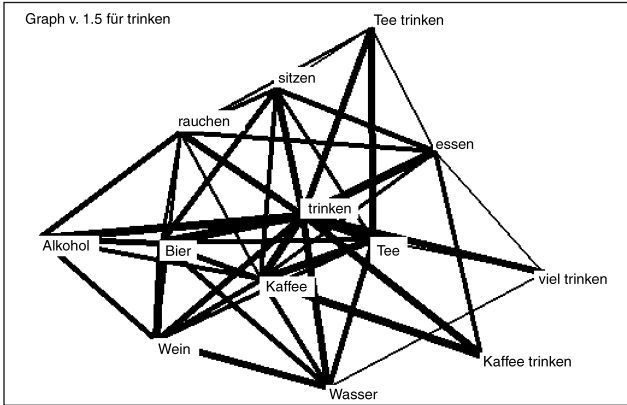


Abb. 7: Netzbasierte Darstellung der Kookkurrenzen zu *trinken* im „Deutschen Wortschatz“

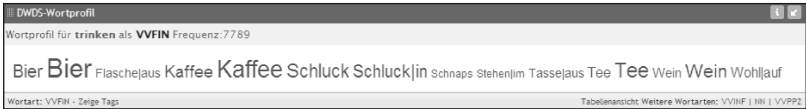


Abb. 8: Typografische Darstellung der Kookkurrenzen zu *trinken* im „DWDS“

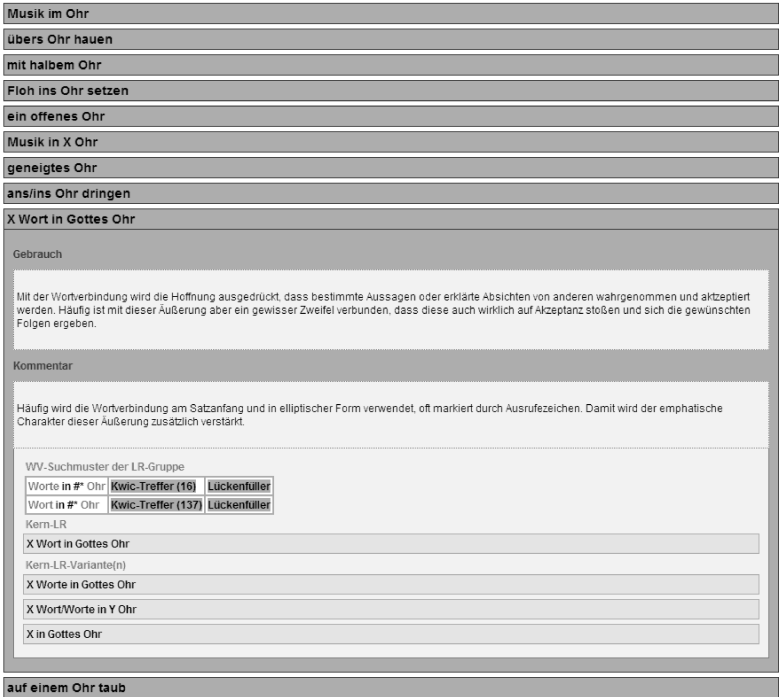


Abb. 9: Ausschnitt aus dem Wortverbindungs-feld *Ohr* in „Wortverbindungen online“

Füller für Suchmuster			
Feld 1	Feld 2	Feld 3	Feld 4
Wort	in	#*	Ohr

Die Füller können durch Klicken auf den Namen der Tabellenspalten sortiert werden.

Füllertabelle für Feld 3 (#*)

Lückenfüller	Anzahl	Prozentanteil
Menschen	1	0,73
der Amerikaner	1	0,73
(Fußball)-Gottes	1	0,73
des Wettergottes	1	0,73
des Theatergottes	1	0,73
Kabilas	1	0,73
Tschubais	1	0,73
Orffens	1	0,73
Wachters	1	0,73
der heimischen Pistenartisten	1	0,73
des Kanzlers	1	0,73
Gottes	73	53,28
Luizaos	1	0,73
Diebes	1	0,73
Levys	1	0,73
der Rechtsgelehrten	1	0,73
Bacchus	1	0,73
des Täters	1	0,73
Konsels und Prohaska	1	0,73
des Sachbearbeiters	1	0,73
des Fußball-Gottes	2	1,46
Handys	1	0,73

Abb. 10: Füller zu der Wendung *X Wort in [Gottes] Ohr* in „Wortverbindungsfelder online“

Doch auch in diesem Bereich gibt es noch viele Lücken zu füllen. Dazu gehört ein großes lexikografisch bearbeitetes Internetkollokationswörterbuch des Deutschen. Auch ein beschreibendes Werk „klassischer Konstruktionen“, also schwer regelhaft herleitbarer, partiell lexikalisch gefüllter Ausdrücke wie etwa die in diesem Band besprochenen *Was-in-aller-Welt*-Konstruktion, *X-hat-seinen-Preis*-Konstruktion (vgl. Stefanowitsch in diesem Band), die verblosen Direktiva und sequenziellen Nominalreduktionen (vgl. Müller in diesem Band) ist noch nicht erarbeitet.

Die Zugänglichkeit zu lexikografischen Daten wird sich aber nicht nur durch die Entwicklung spezifischer lexikografischer Produkte verbessern, sondern auch dadurch, dass bestehende lexikografische Ressourcen zusammengeführt werden. Das kann in Form von Projekten geschehen, die ganz spezifische lexikografische Ressourcen aufeinander abbilden, wie zum Beispiel paradigmatisch organisierte WordNets auf syntagmatisch ausgerichtete FrameNets (siehe Fellbaum in diesem Band). Die Zusammenführung kann aber auch durch die Implementierung von Wörterbuchportalen erfolgen.

Wörterbuchportale integrieren eigenständige sprachlexikografische Referenzwerke, die im Falle von Wörterbuchnetzen wie etwa „OWID“ (Projekt 16) wörterbuchübergreifende Verlinkungen und Zugriffsmöglichkeiten bieten (vgl. Müller-Spitzer/Engelberg, ersch.). Auch aus diesen Bereichen sind in Zukunft für den Bereich der Syntagmatik des deutschen Wortschatzes wichtige Impulse sowohl für die wortschatzbezogene Forschung als auch für die lexikografische Praxis zu erwarten.

Literatur

- Abel, Andrea (2008): ELDIT – Elektronisches Lernerwörterbuch Deutsch–Italienisch und *lexiko*: ein Vergleich. In: Klosa (Hg.), S. 175–189. Internet: www.ids-mannheim.de/pub/laufend/opal/privat/pdf/opal08-1_abel.pdf (Stand: 12.10.2010).
- Abel, Andrea et al. (2005): ELDIT – Elektronisches Lernerwörterbuch Deutsch–Italienisch: einige Aspekte der syntagmatischen und paradigmatischen Ebene. In: Di Meola, Claudio/Hornung, Antonie/Rega, Lorenza (Hg.): Perspektiven Eins. Akten der ersten Tagung Deutsche Sprachwissenschaft in Italien (Rom, 6.–7. Februar 2004). (= Italienische Studien zur deutschen Sprache 1). Rom, S. 589–606.
- Boas, Hans C. (Hg.) (2009): Multilingual FrameNets in computational lexicography: methods and applications. (= Trends in Linguistics: Studies and Monographs 200). Berlin/New York.
- Burchardt, Aloscha et al. (2009): FrameNet for the semantic analysis of German: Annotation, representation and automation. In: Boas (Hg.), S. 209–244.
- Engelberg, Stefan (2010): Die lexikographische Behandlung von Argumentstrukturvarianten in Valenz- und Lernerwörterbüchern. In: Fischer, Klaus/Fobbe, Eilika/Schierholz, Stefan J. (Hg.): Valenz und Deutsch als Fremdsprache. (= Deutsche Sprachwissenschaft international 6). Frankfurt a.M., S. 113–141.
- Engelberg, Stefan/Klosa, Annette/Müller-Spitzer, Carolin (2009): Internet lexicography as a challenge: The Internet dictionary portal at the Institute for German Language. In: Kernerman Dictionary News 17, S. 16–25.
- Gries, Stefan T./Stefanowitsch, Anatol (2010): Cluster analysis and the identification of collexeme classes. In: Newman, John/Rice, Sally (Hg.): Empirical and experimental methods in cognitive/functional research. Stanford, S. 59–71.
- Herbst, Thomas et al. (2004): A valency dictionary of English. A corpus-based analysis of the complementation patterns of English verbs, nouns and adjectives. (= Topics in English Linguistics 40). Berlin/New York.
- Klappenbach, Ruth/Steinitz, Wolfgang (Hg.) (1967–1977): Wörterbuch der deutschen Gegenwartssprache. Bd. 1–6: A-Zytologie. Berlin.
- Klein, Wolfgang (2004): Das Digitale Wörterbuch der deutschen Sprache des 20. Jahrhunderts (DWDS). In: Scharnhorst, Jürgen (Hg.): Sprachkultur und Lexikographie. Von der Forschung zur Nutzung von Wörterbüchern. (= Sprache 50). Frankfurt a.M., S. 281–309.

- Klosa, Annette (2007): Korpusgestützte Lexikographie: besser, schneller, umfangreicher? In: Kallmeyer, Werner/Zifonun, Gisela (Hg.): Sprachkorpora – Datenmengen und Erkenntnisfortschritt. (= Jahrbuch des Instituts für Deutsche Sprache 2006). Berlin/New York, S. 105–122.
- Klosa, Annette (Hg.) (2008): Lexikografische Portale im Internet. (= OPAL – Online publizierte Arbeiten zur Linguistik 1/2008). Mannheim. Internet: www.ids-mannheim.de/pub/laufend/opal/privat/opal08-1.html (Stand: 12.10.2010).
- Kubczak, Jacqueline (2009): Hier wird Ihnen geholfen! E-VALBU – Das elektronische Valenzwörterbuch deutscher Verben. In: Sprachreport 4/2009, S. 17–23.
- Kunze, Claudia/Lemnitzer, Lothar (2010): Lexical-semantic and conceptual relations in GermaNet. In: Storjohann, Petra (Hg.): Lexical-semantic relations. Theoretical and practical perspectives. (= *Linguisticae investigationes: Supplementa* 28). Amsterdam/Philadelphia, S. 163–183.
- Langanke, Ulrich (2004): Das Hypermedia-Online-Wörterbuch an der Schnittstelle zwischen Philologie, Kognition und Informatik. Am Beispiel des „Wortschatz-Lexikons“ (<http://www.wortschatz.uni-leipzig.de>). In: Brdar-Szabó, Rita/Knipp-Komlósi, Elisabeth (Hg.): Lexikalische Semantik, Phraseologie und Lexikographie. Abgründe und Brücken. Festgabe für Regina Hessky. (= Duisburger Arbeiten zur Sprach- und Kulturwissenschaft 57). Frankfurt a.M. u.a., S. 379–393.
- Moerdijk, Fons (2008): Das „Algemeen Nederlandse Woordenboek“ (ANW) und *lexiko* – ein Vergleich. In: Klosa (Hg.), S. 143–151. Internet: www.ids-mannheim.de/pub/laufend/opal/privat/pdf/opal08-1_moerdijk.pdf (Stand: 12.10.2010).
- Moerdijk, Fons/Tiberius, Carole/Niestadt, Jan (2008): Accessing the ANW dictionary. In: Zock, Michael/Huang, Chu-Ren (Hg.): *Coling 2008: Proceedings of the workshop on Cognitive Aspects on the Lexicon (COGALEX 2008)*, Manchester, August 2008, S. 18–24. Internet: www.aclweb.org/anthology/W08-1903 (Stand: 13.10.2010).
- Müller-Spitzer, Carolin (2010): OWID – A dictionary net for corpus-based lexicography of contemporary german. In: Dykstra, Anne/Schoonheim, Tanneke (Hg.): *Proceedings of the XIVth EURALEX International Congress* (Leeuwarden, 6–10 July 2010). Leeuwarden, S. 445–452.
- Müller-Spitzer, Carolin/Engelberg, Stefan (ersch.): Dictionary portals. In: Gouws, Rufus H. et al. (Hg.): *Dictionaries. An international encyclopedia of lexicography. Supplementary volume: Recent developments with special focus on computational lexicography*. Berlin/New York.
- Müller-Spitzer, Carolin/Vachková, Marie (2010): Projektmesse „Elektronische Lexikographie“. In: Sprachreport 2/2010, S. 20–23.
- Prinsloo, Daniel J. (2009): The role of corpora in future dictionaries. In: Nielsen, Sandro/Tarp, Sven (Hg.): *Lexicography in the 21st century. In honour of Henning Bergenholtz*. (= *Terminology and Lexicography Research and Practice* 12). Amsterdam/Philadelphia, S. 181–206.
- Ruppenhofer, Josef et al. (2010): FrameNet II: extended theory and practice. Internet: http://framenet.icsi.berkeley.edu/index.php?option=com_wrapper&Itemid=126 (Stand: 13.08.2010).

- Schmidt, Thomas (2009): The Kicktionary – A multilingual lexical resource of football language. In: Boas (Hg.), S. 101–134.
- Schumacher, Helmut et al. (2004): VALBU – Valenzwörterbuch deutscher Verben. (= Studien zur Deutschen Sprache 31). Tübingen.
- Steyer, Kathrin (2009): Zwischen theoretischer Modellierung und praxisnaher Anwendung. Zur korpusgesteuerten Beschreibung usueller Wortverbindungen. In: Mella-do Blanco, Carmen (Hg.): Theorie und Praxis der idiomatischen Wörterbücher. (= Lexicographica, Series maior 135). Tübingen, S. 119–145.
- Storjohann, Petra/Möhrrs, Christine (2007): Das *elexiko*-Wörterbuch im Wandel – Typische Verwendungen neu beleuchtet. In: Sprachreport 3/2007, S. 12–16.
- Vachková, Marie/Belica, Cyril (2009): Self-organizing lexical feature maps. Semiotic interpretation and possible application in lexicography. In: Interdisciplinary Journal for Germanic Linguistics and Semiotic Analysis 13, 2, S. 223–260.

Links zu den vorgestellten Projekten

- Projekt 1 (*Algemeen Nederlands Wordenboek*): <http://anw.inl.nl/search>
- Projekt 2 (*Das Große akademische Wörterbuch Deutsch-Tschechisch*): https://lexarchiv.ff.cuni.cz/lexikograficka-sekce/_spravce/
- Projekt 3 (*Deutscher Wortschatz der Universität Leipzig*): <http://wortschatz.uni-leipzig.de/>
- Projekt 4 (*Digitales Wörterbuch zur deutschen Sprache – DWDS*): www.dwds.de
- Projekt 5 (*Entwicklung eines korpusbasierten, elektronischen Wörterbuchs Deutsche Gebärdensprache – Deutsch*): www.sign-lang.uni-hamburg.de/dgs-korpus/
- Projekt 6 (*GermaNet – ein lexikalisch semantisches Wortnetz*): www.sfs.uni-tuebingen.de/lcd/
- Projekt 7 (*HyTex – Werkzeuge zur Exploration von GermaNet*): www.hytex.tu-dortmund.de/ressourcen.html#wortnetze
- Projekt 8 (*Saarbrücken Lexical Semantics Acquisition Project – SALSa*): www.coli.uni-saarland.de/projects/salsa
- Projekt 9 (*Kicktionary*): www.kicktionary.de
- Projekt 11 (*E-VALBU*): <http://hypermedia2.ids-mannheim.de/evalbu/index.html>
- Projekt 12 (*Erlanger Valency Patternbank*): www.patternbank.uni-erlangen.de
- Projekt 13a (*Usuelle Wortverbindungen*): <http://wvonline.ids-mannheim.de/>
- Projekt 13b (*SprichWort*): www.sprichwort-plattform.org/
- Projekt 14 (*elexiko*): www.elexiko.de
- Projekt 15 (*Benutzeradaptive Zugänge und Vernetzungen in elexiko*): www.ids-mannheim.de/lexik/BZVelexiko/
- Projekt 16 (*Online-Wortschatz-Informationssystem Deutsch – OWID*): www.owid.de